

# cDNA structure, expression and nucleic acid-binding properties of three RNA-binding proteins in tobacco: occurrence of tissue-specific alternative splicing

Tetsuro Hirose, Mamoru Sugita and Masahiro Sugiura\*  
Center for Gene Research, Nagoya University, Nagoya 464-01, Japan

Received June 7, 1993; Accepted July 14, 1993

DDBJ accession nos D16204–D16206 (incl.)

## ABSTRACT

Three cDNAs encoding RNA-binding proteins were isolated from a tobacco (*Nicotiana glauca*) cDNA library. The predicted proteins (RGP-1) are homologous to each other and consist of a consensus-sequence type RNA-binding domain of 80 amino acids in the N-terminal half and a glycine-rich domain of 61–78 amino acids in the C-terminal half. Nucleic acid-binding assay using the *in vitro* synthesized RGP-1 protein confirmed that it is an RNA-binding protein. Based on its strong affinity for poly(G) and poly(U), the RGP-1 proteins are suggested to bind specifically to G and/or U rich sequences. All three genes are expressed in leaves, roots, flowers and cultured cells, however, the substantial amount of pre-mRNAs are accumulated especially in roots. Sequence analysis and ribonuclease protection assay indicated that significant amounts of alternatively spliced mRNAs, which are produced by differential selection of 5' splice sites, are also present in various tissues. Tissue-specific alternative splicing was found in two of the three genes. The alternatively spliced mRNAs are also detected in polysomal fractions and are suggested to produce truncated polypeptides. A possible role of this alternative splicing is discussed.

## INTRODUCTION

The gene expression in eukaryotic cells is known to be regulated at various levels. The regulation at the post-transcriptional level (e.g. at the steps involving RNA capping, RNA splicing, RNA polyadenylation, RNA transport and/or RNA stability) has recently been recognized as a critical factor for the expression of several genes in animal and fungal cells (1, 2). These events are complex and require many proteins and/or RNA factors. The heterogeneous nuclear ribonucleoprotein (hnRNP) complexes, which are composed of pre-mRNAs, 20–25 different proteins and small nuclear ribonucleoprotein (snRNP) complexes (consisting of U-snRNAs and multiple proteins), are involved in nuclear pre-mRNA processing and splicing (3). Many RNA-binding proteins related to these steps contain one or more copies

of a conserved domain, named as, consensus-sequence type RNA-binding domain (CS-RBD) of about 80 amino acids (4). CS-RBD includes two highly conserved motifs, one an octamer, termed as ribonucleoprotein consensus-sequence (RNP-CS) and, another a hexamer, RNP-2 (5). CS-RBD has been shown to be the minimum structure for RNA-binding activity (6).

We have isolated five RNA-binding proteins (or ribonucleoproteins, RNPs) from tobacco chloroplasts (7, 8). Analysis of their cDNAs and genes revealed that these proteins were nuclear encoded and contain two typical CS-RBDs and an acidic amino terminal domain (7–10). Nucleic acid-binding assay of these proteins revealed their preferential binding to poly(G) and poly(U) (11, 12). Recently similar proteins and their cDNAs were isolated from spinach and maize chloroplasts (13, 14) and the spinach 28RNP has been shown to be required for the pre-mRNA processing from the chloroplast *petD* gene (14). Two cDNAs encoding chloroplast RNA-binding proteins were also isolated from *Nicotiana plumbaginifolia* using two sets of oligonucleotides designed on the basis of RNP-CS and RNP-2 (15).

In plants, except for these chloroplast RNPs, little is known about protein factors related to various nuclear RNA processing events. A maize gene whose expression was induced by abscisic acid and water stress was isolated and reported to encode a glycine-rich protein (16). The deduced protein was later pointed out to contain a sequence which conforms to RNP-CS (17). Recently the homologous cDNAs have been reported from different plant species; e.g. a maize cDNA induced by heavy metal shock (18), two cDNAs in sorghum (19) and one carrot cDNA induced by wounding (20). All of them contain CS-RBD-like sequences, although, their nucleic acid-binding properties have not been characterized. We have attempted to isolate more cDNAs encoding consensus-sequence type RNA-binding proteins from tobacco using oligonucleotide probes, designed by using RNP-CS. Here we present three cDNAs which encode consensus-sequence type RNA-binding proteins with high affinities for poly(U) and poly(G) but do not encode the transit peptides in their N-termini. All the three genes are expressed in leaves as well as in roots while substantial levels of alternatively spliced mRNAs are present in various tissues. Moreover, two of the three genes also display alternative splicing in tissue-specific manner.

\* To whom correspondence should be addressed

## MATERIALS AND METHODS

### cDNA isolation and sequencing

An oligonucleotide probe (RNP-1) was prepared based on the amino acid sequences of RNP-CSs of chloroplast ribonucleoproteins as shown below.

Amino acid sequence of RNP-CS and flanking residues of two each:

```

  T G R S R G F G F V T M S
  D                               F

```

Deduced nucleotide sequence:

```

5' ACX GGX ACX AGX CGX GGX TTT GGX TTT GTX ACX ATX TC 3'
   A C  TC T          CC  CA      T

```

RNP-1 probe:

```

3' TGI CCI TCI TCI GCI CCI AAA CCI AAA CAI TGI AAI AG 5'
   T G  AG A          GG   GT   T

```

Screening of a tobacco (*Nicotiana glauca*)  $\lambda$ gt10 cDNA library was essentially according to the instruction manual of Hybond-N membrane (Amersham). The oligonucleotide probe was 5' end-labeled by polynucleotide kinase and hybridizations were carried out at 50°C. Positive clones were isolated and the sizes of the inserts were estimated by PCR method described by Li and Sugiura (7). Recombinant IDNAs were prepared by the plate lysate method (21). The inserts were then cut with *EcoRI*, subcloned into pBluescript SK<sup>+</sup> and sequenced by dideoxy chain termination method (22).

### *In vitro* transcription and translation

The DNA fragment encoding RGP-1b was prepared by PCR using its cloned cDNA as a template and primers as below:

```

exp1b = 5' GCTCGAGCTGAAGTAGAATACAGTTGC 3'
reverse = 5' AACAGCTATGACCATG 3'

```

The amplified fragment was cut with *XhoI/EcoRI*, purified by 1% agarose gel electrophoresis and ligated after the TMV leader of TMV expression vector at the *XhoI* site (11). The sequence encoding RGP-1b was verified by DNA sequencing. A capped transcript was produced using a mRNA capping kit (mCAP<sup>TM</sup> kit, Stratagene) and translation was performed in a rabbit reticulocyte *in vitro* translation system (Promega) as described previously (12). The N-terminal sequence of *in vitro* synthesized protein is MARAEVE-.

### Nucleic acid-binding assay

Binding assay was carried out essentially according to Ye and Sugiura (12). The *in vitro* synthesized protein labeled with <sup>35</sup>S was mixed with 20  $\mu$ l each of ssDNA-cellulose (0.75–1.5 mg DNA/ml), dsDNA-cellulose (0.75–2 mg DNA/ml), tobacco total RNA-Sepharose (0.6 mg RNA/ml), and four kinds of ribonucleotide homopolymer-Sepharose (poly(G): 0.21 mg/ml, poly(A): 0.8 mg/ml, poly(U): 0.22 mg/ml, poly(C): 0.24 mg/ml) in 1ml buffer B (10 mM Tris-HCl, pH 7.6, 2.5 mM MgCl<sub>2</sub>, 0.5% Triton X-100, 1  $\mu$ g/ml pepstatin, 1 mM PMSF and 0.1 to 2.0 M NaCl). The mixture was incubated at 4°C for 10 min with gentle shaking. The beads were washed successively once with buffer B containing 2mg/ml heparin, twice with binding buffer B and twice with distilled water. Bound proteins were eluted with the loading buffer for SDS-polyacrylamide gel electrophoresis (PAGE) and 15  $\mu$ l of released proteins were

applied to 0.1% SDS/12.5% polyacrylamide gels. After electrophoresis at 100 V for 4–5 h, the gels were dried and exposed to an imaging plate (Fuji Photo film Co) overnight. The relative amount of bound proteins was calculated by a Fuji Bioimage analyzer BAS2000 (Fuji Photo Film Co.).

### Total RNA isolation and Northern analysis

Isolation of total RNA from *N. sylvestris* leaves, roots and flowers was performed by hot phenol method (23). Total RNA from tobacco (*N. tabacum* BY-2) cultured cells was obtained by the small scale preparation method using mini-BeadBeater (24). RNA electrophoresis and Northern blotting to Hybond-N membrane were according to the instruction manual (Amersham). Oligonucleotide probes (UTR-1a, -1b, -1c; sequences and positions are shown below) corresponding to the 3' untranslated regions of the three cDNAs were labeled at their 5' ends using polynucleotide kinase (specific activity of 2.8–5.1  $\times 10^6$  cpm/pmol).

UTR-1a (positions 896-937 in the RGP-1a cDNA sequence).

5' CCACAGTAAACCATAACGGAACCTCAACCAAACCTTAGAACCA 3'

UTR-1b (positions 777–820 in the RGP-1b cDNA sequence).

5' ACCAACCACACTAAAACAGTAATGGATACTAATCTAAAGGCCAA 3'

TR-1c (positions 875–914 in the RGP-1c cDNA sequence).

5' GCGATCAAATAACTAAAATCCACATCTTCTCAATTATCT 3'

Hybridization and washing were done according to the oligonucleotide protocol of ZETA-PROBE instruction manual (Bio-Rad). Hybridized and washed membranes were exposed to an imaging plate for 48 h and analyzed by a Fuji Bioimage analyzer BAS2000.

### Polysomal RNA preparation

Polysomal RNA was prepared according to de Vries et al. (25). About 60 g of *N. sylvestris* young leaves were homogenized in liquid nitrogen by a motor and pestle. The resultant leaf powder was gently mixed with 180 ml of polysome buffer (50 mM Tris-HCl, pH 9.0, 50 mM MgCl<sub>2</sub>, 25 mM EGTA, 250 mM NaCl, 1% Nonidet P40) and gently ground until ice pieces completely disappeared. This suspension was centrifuged at 12000 rpm for 15 min and the supernatant was filtered through single layer of Whatman 3MM paper. The filtrate was overlaid onto 60% sucrose cushion and centrifuged at 40000 rpm for 3 h at 2°C. The polysome/ribosome pellets were suspended in 3 ml of gradient buffer (10 mM Tris-HCl, pH 8.5, 10 mM MgCl<sub>2</sub>, 5 mM EGTA, 50 mM NaCl) and overlaid on 10–40% sucrose gradients and centrifuged at 25000 rpm for 70 min at 2°C. After fractionation, polysomal fractions were collected and precipitated with ethanol. RNA was extracted as described above.

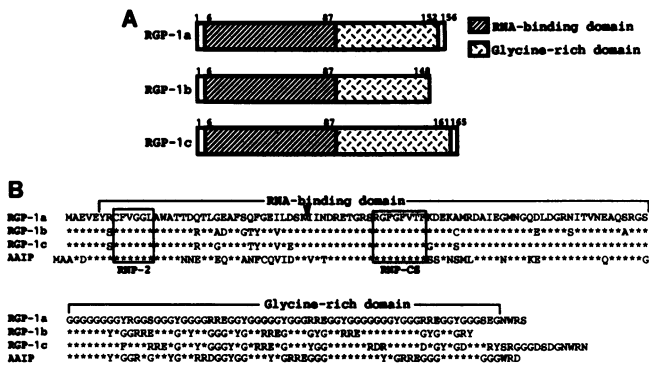
### Ribonuclease protection assay

This was carried out according to the instruction manual of RPAII kit (Ambion). Plasmid DNAs were linearized with *EcoRI*. Radioactive antisense RNA probes were synthesized using an *in vitro* transcription kit (Stratagene), and purified by 6% PAGE in the presence of 7 M urea. RNAs from various tissues were treated with RNase-free DNase I to remove trace amounts of contaminating DNA. RNA probes (about 1  $\times 10^5$  cpm) were hybridized overnight to 10  $\mu$ g of tissue RNA at 42°C. The hybridized RNA was digested with a mixture of 0.1 units of RNaseA and 100 units of RNase T1 at 37°C for 30 min. The protected fragments were separated on 6% polyacrylamide gels containing 7 M urea and analyzed by a Fuji Bioimage analyzer BAS2000.

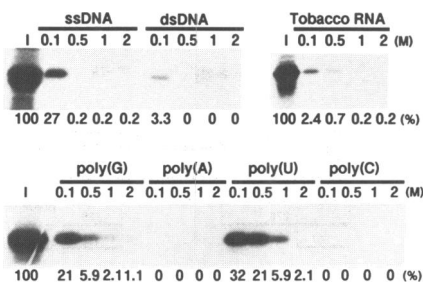
**RESULTS**

**Structure of the cDNA encoding proteins containing CS-RBDs**

An oligonucleotide probe (RNP-1), complementary to the deduced nucleotide sequence from the RNP-CS amino acid sequence (7, 8) was used to screen a *Nicotiana sylvestris* leaf cDNA library. Out of eleven positive clones isolated, cDNA inserts of six clones were completely sequenced, whereas those of the remainder were partially sequenced. These clones could be distinguished four groups according to their predicted protein structures. The cDNAs of three groups (about 0.7–1.0 kb in length) encode proteins consisting of a single CS-RBD in the N-terminal half and a glycine-rich domain in the C-terminal half (Fig. 1A), which were named as RNA-binding Glycine-rich Protein-1 (RGP-1) (RGP-1a, 1b, 1c). They do not seem to be imported into organelles as no apparent transit peptides could be detected. The fourth group is distinct from RGP-1 group as it has only 50% homology and was named as RGP-2 and not analyzed further. The three predicted RGP-1 proteins are highly homologous to each other (1a:1b=81%, 1a:1c=80%, 1b:1c=82%) (Fig. 1B).



**Figure 1.** Structure of the RGP-1 proteins deduced from their cDNA sequences. A. Domain structures of the three RGP-1 proteins. Numbers indicate amino acid positions from the first methionine. B. Alignment of the three RGP-1 protein sequences and the abscisic acid induced glycine-rich protein (AAIP) sequence in maize (16). Asterisks represent identical amino acids to that of RGP-1a. An arrowhead indicates the insertion site of additional sequences (see Fig. 3). The sequences are in DDBJ/EMBL/GenBank accession nos. D16204 (RGP-1a), D16205 (RGP-1b), D16206 (RGP-1c).

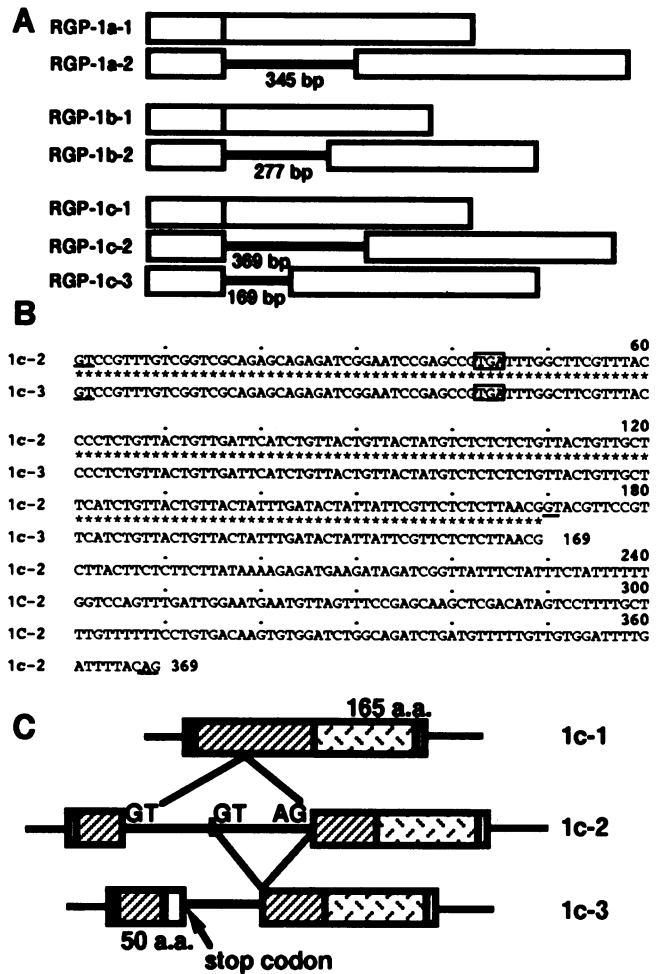


**Figure 2.** Nucleic acid-binding properties of the RGP-1b protein. The *in vitro* synthesized [<sup>35</sup>S] protein was mixed with various nucleic acids, ssDNA and dsDNA (calf thymus), total RNA (*N. sylvestris*), and four kinds of ribonucleotide homopolymers, at indicated salt concentrations (M). Bound proteins were analyzed by SDS-PAGE. Numbers under the respective lanes indicate the relative amount (%) of bound proteins to that of the input protein (lanes 1).

The RNP-CS (RGFGFVTF) and RNP-2 (CFVGGL) sequences in their CS-RBDs are identical among the RGP-1 group. About 60% of glycine residues in the glycine-rich domain are contiguous, and single tyrosines and charged amino acid stretches (RRE or RDR) are inserted between the glycine stretches. The protein-coding regions of the cDNAs are highly conserved, however the homology of respective 5' and 3' untranslated regions is relatively low. Each of the RGP-1 proteins showed about 70% homology with the deduced amino acid sequence of an abscisic acid induced cDNA encoding glycine-rich protein in maize (16).

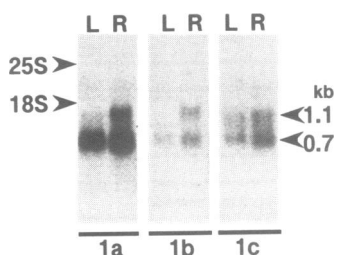
**Nucleic acid-binding property of the RGP-1b protein**

To examine its affinity to nucleic acids, the RGP-1b protein was synthesized *in vitro* and its binding to DNAs and RNAs was assayed. Labeled RGP-1b protein was incubated with nucleic



**Figure 3.** Additional sequences in the cDNAs. A. Sequenced cDNAs. RGP-1a-1 and RGP-1c-1 were derived from PCR amplification using respective specific primers corresponding to the 3' untranslated regions and the N-terminal protein coding regions. Open boxes represent exons from start codons to stop codons and lines with bp show additional sequences. B. Alignment of the nucleotide sequences of the additional sequences in RGP-1c-2 and RGP-1c-3. Asterisks indicate identical nucleotides. Bordering sequence between exons and introns, GT and AG, are underlined. The first stop codon in exon reading frames is boxed. C. Schematic model of the formation of the three RGP-1c cDNAs. Boxes are as described in Fig. 1A.

acid-cellulose or -Sepharose beads at 0.1–2.0 M NaCl. After washing with buffer containing heparin, the bound proteins were separated by SDS-PAGE. The RGP-1b protein remains bound to both ssDNA- and dsDNA-cellulose up to 0.1 M NaCl and to total tobacco RNA up to 1 M NaCl (Fig. 2). We therefore concluded that the RGP-1b protein (and probably 1a and 1c proteins) is an RNA-binding protein. However, the relative amount of protein bound to ssDNA at 0.1 M NaCl is higher than that bound to total tobacco RNA (Fig. 2). This suggests that the RGP-1 protein binds nonspecifically to nucleic acids at a low salt concentration whereas its binding to a subpopulation of tobacco total RNA is more specific. We next examined the binding properties of their proteins to poly(G), poly(A), poly(U) and poly(C). Fig. 2 shows that the RGP-1b protein binds specifically to poly(G) and poly(U) up to 2 M NaCl, suggesting that it binds specifically to G/U rich RNA species and/or G/U rich regions of RNA molecules.



**Figure 4.** Northern blot analysis of transcripts from the three protein (1a, 1b and 1c) genes. *N.sylvestris* leaf (L) and root (R) RNA (10  $\mu$ g each), and oligonucleotide probes (UTR-1a, -1b, -1c) that are complementary to parts of the 3' untranslated regions were used. Size markers are tobacco 25S and 18S rRNAs.

#### Additional sequences found in some of the cDNAs

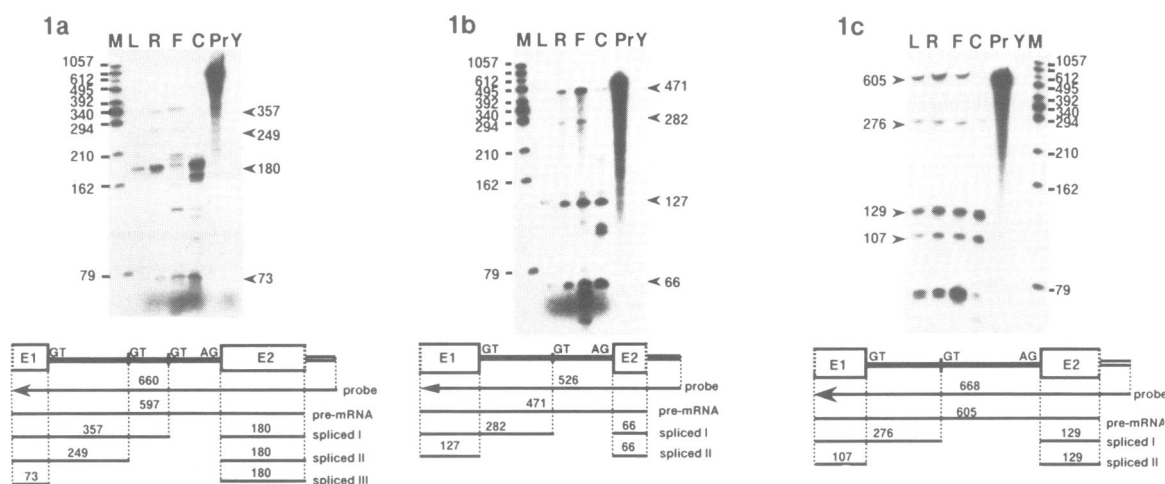
Four cDNAs among the analyzed cDNAs were found to contain an additional sequence of 169–369 bp with respect to the corresponding RGP-1 cDNA (Fig. 3A). The insertion of these sequences is at the same position and is in between the region encoding RNP-2 and RNP-CS in CS-RBD. This site also corresponds to the insertion site of an intron in the genes encoding chloroplast RNA-binding proteins (8–10).

Furthermore these additional sequences contain the GT-AG consensus bordering sequence and A/T rich nucleotide composition typical to the intron sequence of dicot plants (26). Therefore the additional sequence is likely to be an unprocessed intron sequence. Direct sequencing of the PCR fragments (amplified from tobacco genomic DNA using RGP-1c specific primers) also confirmed that the additional sequence matches perfectly to the RGP-1c genomic sequence. These results indicate that the additional sequence is derived from an intron and that these cDNAs were synthesized from pre-mRNAs. Furthermore, it is suggested that these genes have only one intron, at least in the coding region.

Interestingly, the additional sequence with different sizes (169 bp and 369 bp) were found in a RGP-1c population (Fig. 3A). The additional sequence (169 bp) of the RGP-1c-3 cDNA matches perfectly with the sequence 169 bp 5' to the additional sequence (369 bp) of RGP-1c-2 cDNA (Fig. 3B).

#### Transcript levels of the RGP-1 genes

Northern blot analysis of tobacco leaf and root RNAs was performed using oligonucleotide probes corresponding to the 3' untranslated region of the three cDNAs (the 3' untranslated region is less conserved than the coding region). All three genes were found to be transcribed in both tissues but their transcript levels were higher in roots than in leaves (Fig. 4). Interestingly, two different transcripts (0.7 kb and 1.1 kb) from the three genes were detected both in leaves and roots. Based on the size of these



**Figure 5.** Detection of mRNA species from the RGP-1a (1a), RGP-1b (1b) and RGP-1c (1c) genes by ribonuclease protection assay. Bands indicated by arrowheads with length in nt represent protected fragments from 10  $\mu$ g of *N.sylvestris* RNA hybridized with an antisense RNA probe. M, size markers ( $\phi$ x174 RF-DNA digested with *HincII*); L, leaf RNA; R, root RNA; F, flower RNA; C, cultured BY-2 cell RNA; Y, yeast RNA as control; Pr, RNA probe. The experimental design is shown below. Arrows represent antisense RNA probes corresponding to a 3' portion of exon 1 (E1) + an intron (bold line) + a 5' portion of exon 2 (E2) + a 63 nt portion of pBluescript sequence (double line): 1a, RGP-1a-2 (660 nt, positions 74–670 + 63 nt), 1b, RGP-1b-2 (526 nt, positions 1–471 + 55 nt) and 1c, RGP-1c-2 (668 nt, positions 11–615 + 63 nt). Solid lines designate protected fragments with length in nt. Asterisks indicate the first stop codons in introns.

two transcripts, it is suggested that the 1.1 kb RNAs are the pre-mRNAs and 0.7 kb RNAs are the mature mRNAs. Northern blot analysis using an oligonucleotide probe corresponding to a part of the intron also confirmed that only 1.1 kb RNAs are the intron-containing RNAs (data not shown).

Next we analyzed the effect of drought stress and abscisic acid on the gene expression of the RGP-1 genes. Young tobacco plants were either sprayed with 0.1 mM abscisic acid or desiccated by air drying. After 12 and 24 h total RNA was prepared from these plants and analyzed by northern hybridization using the above oligonucleotide probes. However no obvious changes were detected for the transcript levels of each RGP-1 gene (data not shown).

### Pre-mRNA splicing of the RGP-1 genes

As described above, we isolated two different RGP-1c cDNAs which contain the additional sequence of different sizes, 169 bp (the 5' half of the intron) in RGP-1c-3 and 369 bp (the full intron size) in RGP-1c-2 (see Fig. 3). The 169 bp sequence is followed by GT, the 5' intron boundary sequence, in the 369 bp intron (see Fig. 3B). This suggests that the RGP-1c-3 cDNA is derived from an alternatively spliced mRNA, due to alternative selection of the 5' splice site (Fig. 3C). In order to confirm this suggestion, ribonuclease protection assay was performed using an antisense RNA probe synthesized from a part of RGP-1c-2 cDNA (derived from the pre-mRNA) and total RNAs from tobacco tissues. It is observed that pre-mRNA, alternatively spliced RNA and fully spliced RNA are present in all four tissues examined, viz. leaves, roots, flowers and cultured cells (Fig. 5, 1c).

Ribonuclease protection assay was also performed for RGP-1a and 1b transcripts as in the case of RGP-1c. The results showed that alternatively spliced RGP-1a and 1b mRNAs are present, in roots and flowers but not in leaves and cultured cells, indicating that alternative splicing occurs in tissue-specific manner (Fig. 5, 1a and 1b). Based on the size of protected fragments and the positions of potential 5' splice sites (GT) in the introns, in RGP-1a two alternatively spliced RNAs are likely to be present in roots and flowers whereas in the case of RGP-1b only one alternatively spliced RNA species is probably present. The substantial amounts of pre-mRNAs accumulated in RGP-1b (and 1c) but not in RGP-1a. These results indicate that the alternative splicing by differential selection of 5' splice sites generally occurs in all pre-mRNAs from the three RGP-1 genes. These alternatively spliced mRNAs from all three genes might result in truncated polypeptides upon translation (see Fig. 3B, 3C).

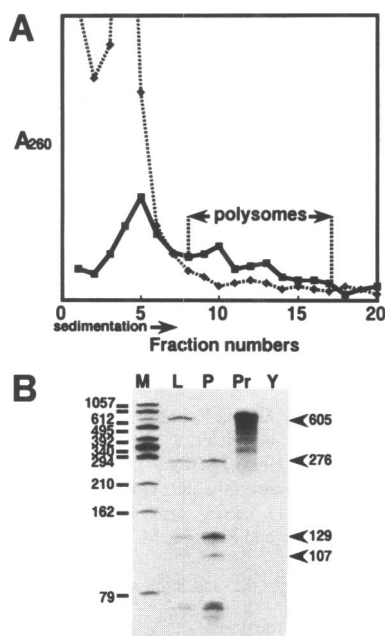
### Detection of the alternatively spliced mRNA in polysomal fractions

To find out whether two kinds of polypeptides are in fact produced due to alternative splicing of the RGP-1c pre-mRNA, we prepared polysomal fractions from tobacco young leaves and analyzed the RGP-1c mRNA species in the polysomal RNA pool by ribonuclease protection assay (Fig. 6A). Both fully and alternatively spliced mRNAs were found in the polysomal fraction while the pre-mRNA could not be detected (Fig. 6B). This result clearly indicates that the alternatively spliced mRNA is transported from the nucleus to the cytoplasm and incorporated into polysomes. When the alternatively spliced mRNA is translated, a termination codon emerges next to the fourteenth codon from the distal 5' splicing site and a short polypeptide of 50 amino acids is expected to be produced (see Fig. 3C).

### DISCUSSION

We have isolated three related cDNAs encoding consensus-sequence type RNA-binding proteins from *N. sylvestris*. These predicted proteins include a single CS-RBD towards N-terminal, a glycine-rich domain towards C-terminal half but no transit peptide (Fig. 1A). They are homologous to each other and also to a maize protein induced by abscisic acid (16) (Fig. 1B). CS-RBDs are highly conserved, however, homology among glycine-rich domains is lower than that of CS-RBDs. Repetitive units like GGGYGGG and GGGRREGGG are commonly present in the glycine-rich domain of these proteins. Glycine stretches with tyrosines are also found in glycine-rich proteins lacking CS-RBD in tobacco and *Arabidopsis* (27, 28). A glycine-rich domain is found in the several RNA-binding proteins, in the C-terminal region of animal hnRNPs A1 (29) and A2/B1 (30), in the inner region of animal nucleolin (31), the spacer region of the cp29 protein from tobacco chloroplasts (8). In the hnRNP A1 protein, this domain has been shown to enhance its affinity to RNA (32). Therefore, the glycine-rich domain in the RGP-1 proteins might also function as in the case of hnRNP A1 protein.

Our nucleic acid-binding assay using the *in vitro* synthesized RGP-1b protein has confirmed that it really is an RNA-binding protein, however, the strength of binding to each nucleic acid is different. At a low salt concentration the quantity of bound protein was larger to ssDNA than to tobacco total RNA, whereas at a higher salt concentration it bound specifically to RNA (Fig. 2). This suggests that this protein binds to specific sequences



**Figure 6.** Detection of the RGP-1c mRNAs in polysomes. **A.** Fractionation of polysomes/ribosomes from *N. sylvestris* leaves by sucrose density gradient in the presence of 10 mM MgCl<sub>2</sub> (solid line) and 200 mM EDTA (dashed line). Fractions 8 to 17 were collected and polysomal RNA was extracted. **B.** Ribonuclease protection assay. Bands indicated by arrowheads with length in nt represent protected fragments from 10 µg of *N. sylvestris* polysomal RNA (P) hybridized with the RGP-1c-2 antisense RNA probe. Other details are as in the legend to Fig. 5.

(so a minor fraction) in a total RNA population, while it weakly binds to ssDNA in a nonspecific manner *in vitro*. This prediction is also supported by the result that the RGP-1b protein specifically binds to poly(G) and poly(U).

Based on the Northern blot analysis and ribonuclease protection assay these proteins appear to be expressed constitutively in leaves, roots, flowers and cultured cells (Figs. 4, 5). The fact that the transcript levels of tobacco RGP-1 genes were not changed by drought stress and abscisic acid, supports their constitutive expression. The maize gene and the recently isolated carrot cDNA, encoding structurally related proteins to RGP-1, were reported not to be expressed in leaves and also that their expression was induced by drought, abscisic acid and wounding (16, 20). Therefore, the three RGP-1 genes are distinct from the maize and carrot genes.

The pre-mRNA from the RGP-1c gene was shown to be spliced into two forms in all tissues examined. This alternative splicing results due to the differential selection of two 5' splice sites (Fig. 3C). The pre-mRNA and the two forms of differentially spliced mRNAs are accumulated substantially in leaves, roots and flowers, however, the pre-mRNA was not detected in tobacco BY-2 cultured cells (Fig. 5). These data indicate that efficiency of pre-mRNA splicing is relatively low except in the cultured cells and alternative selection of 5' splice site generally occurs. In RGP-1a and 1b transcripts alternatively spliced mRNAs were detected in roots and flowers but not in leaves and cultured cells, indicating that the alternative splicing operates tissue-specifically. To our knowledge, this is the first report of tissue-specific alternative splicing in plants. These observations suggest that the expression of the RGP-1c gene is critically regulated at the level of pre-mRNA splicing. Alternative splicing in plants has been reported for the RuBisCO activase gene in spinach and *Arabidopsis* (33) and the myb-like P-protein gene in maize (34), both of which produce two functional isoforms. However the putative product of alternative splicing in case of RGP-1c pre-mRNA is truncated because a termination codon appears at the fourteenth amino acid from the distal 5' splice site. Hence, this truncated polypeptide (50 amino acids) does not contain a complete CS-RBD and it unlikely binds to RNA.

We showed that the alternatively spliced mRNAs as well as the fully spliced mRNA are incorporated into polysomes but the pre-mRNA is not (Fig. 6B). Therefore, both the spliced mRNAs are translatable but the pre-mRNAs is not transported into the cytoplasm, hence, not translatable. In addition, this result indicates that the *cis*-element to prevent transport of the pre-mRNA from the nucleus to the cytoplasm resides in the 3' half of the intron sequence. From these cytoplasmic mRNAs not only a functional RNA-binding protein but also a truncated polypeptide is expected to be produced in various tobacco organs. The biological significance of this and the function of the truncated polypeptide is not known yet.

A similar regulatory process is known in the case of the pre-mRNA splicing of several sex determining genes at the developmental stage of sex determination in *Drosophila*. The pre-mRNAs of the sex determination genes (e.g. *sxl*, *dsx*, *tra*) are alternatively spliced by the differential selection of 3' splice sites, resulted in exon skipping at restricted developmental stages (35). These genes (except for *tra*) also encode RNA-binding proteins and produce functional RNA-binding proteins and truncated polypeptides. The alternative splicing of the RGP-1c pre-mRNA appear to occur constitutively while this is not the case for *Drosophila* sex determination genes. Alternative splicing has been

reported to be an essential process of expression of genes encoding several other RNA-binding proteins including hnRNP A1(36), A2/B1(30) and ASF (37).

Besides the RGP-1c gene two other genes (1a and 1b) are likely to have an intron in the same site and their pre-mRNAs are probably spliced alternatively (Fig. 5), suggesting that all three RGP-1 genes potentially produce truncated polypeptides as well as functional RNA-binding proteins. As the truncated polypeptides are expected to be produced from all three RGP-1 genes and are not functional for RNA-binding, the alternative splicing may provide an inhibition mechanism of the expression of functional RNA-binding proteins and the ratio of the usage of two to three 5' splice sites may change under particular condition(s).

The function of RGP-1 proteins remains to be analyzed. However, the structure and the nucleic acid-binding properties of the proteins are similar to those of human hnRNP A1 protein, suggesting that the RGP-1 protein functions in plant cells as hnRNP A1s do in mammalian cells. A recent report indicates that the A1 protein is involved in the 5' splice site selection during alternative splicing in animal cells (38). This fact raises the possibility that the RGP-1 gene expression is self-regulated. The molecular mechanism of various post-transcriptional regulations in plants has been reported to be different at least in part from those in animals and fungi. Further analysis about the function of RGP proteins may give a clue to the elucidation of a novel molecular mechanism in post-transcriptional regulation.

## ACKNOWLEDGMENTS

We thank Dr Y.Li for providing the cDNA library from *N.sylvestris* leaves, Dr L.Ye for preliminary screening of cDNA clones, Dr S.Kapoor for critical reading of this manuscript, Drs T.Wakasugi and A.Vera for their valuable discussions and encouragements. This work was supported in part by a Grant-in-Aid from the Ministry of Education, Science and Culture (Japan).

## REFERENCES

- Higgins, C. F. (1991) *Curr. Opin. Cell Biol.*, **3**, 1013–1018.
- Mattaj, I. W. (1990) *Curr. Opin. Cell Biol.*, **2**, 528–538.
- Bandziulis, R. J., Swanson, M. S. and Dreyfuss, G. (1989) *Genes Dev.*, **3**, 431–437.
- Kenan, D. J., Query, C. C. and Keene, J. D. (1991) *Trends Biochem. Sci.*, **16**, 214–220.
- Query, C. C., Bentley, R. C. and Keene, J. D. (1989) *Cell*, **57**, 89–101.
- Nietfeld, W., Mentzel, H. and Pieler, T. (1990) *EMBO J.*, **9**, 3699–3705.
- Li, Y. and Sugiura, M. (1990) *EMBO J.*, **9**, 3059–3066.
- Ye, L., Li, Y., Fukami-Kobayashi, K., Go, M., Konishi, T., Watanabe, A. and Sugiura, M. (1991) *Nucleic Acids Res.*, **19**, 6485–6490.
- Li, Y., Ye, L., Sugita, M. and Sugiura, M. (1991) *Nucleic Acids Res.* **19**, 2987–2991.
- Li, Y., Nagayoshi, S., Sugita, M. and Sugiura, M. (1993) *Mol. Gen. Genet.* in press.
- Li, Y. and Sugiura, M. (1991) *Nucleic Acids Res.*, **19**, 2893–2896.
- Ye, L. and Sugiura, M. (1992) *Nucleic Acids Res.*, **20**, 6275–6279.
- Cook, W.B. and Walker, J. C. (1992) *Nucleic Acids Res.*, **20**, 359–364.
- Schuster, G. and Grissem, W. (1991) *EMBO J.*, **10**, 1493–1502.
- Mieszczyk, M., Klahre, U., Levy, J. H., Goodall, G. J. and Filipowicz, W. (1993) *Mol. Gen. Genet.*, **234**, 390–400.
- Gomez, J., Sanchez-Martinez, D., Stiefel, V., Rigau, J., Puigdomenech, P. and Pages, M. (1988) *Nature*, **334**, 262–264.
- Mortenson, E. and Dreyfuss, G. (1989) *Nature*, **337**, 312.
- Didierjean, L., Frendo, P. and Burkard, G. (1992) *Plant Mol. Biol.* **18**, 847–849.
- Cretin, C. and Puigdomenech, P. (1990) *Plant Mol. Biol.*, **15**, 783–785.

20. Sturm, A. (1992) *Plant Physiol.*, **99**, 1689–1692.
21. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
22. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
23. Shirzadegan, M., Christie, P. and Seemann, J. R. (1991) *Nucleic Acids Res.*, **19**, 6055.
24. Chang, C. and Meyerowitz, E. M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1408–1412.
25. de Vries, S., Hoge, H. and Bisseling, T. (1988) In Gelvin, S. B., Schilperoort, R. A. and Verma, D. P. S. (ed.), *Plant Mol. Biol. Manual*, Kluwer Academic Publishers, Dordrecht, **B6**, pp. 1–13.
26. Goodall, G. J. and Filipowicz, W. (1991) *EMBO. J.*, **10**, 2635–2644.
27. Obokata, J., Ohme, M. and Hayashida, N. (1991) *Plant Mol. Biol.*, **17**, 953–955.
28. de Oliveira, D. E., Seurinck, J., Inze, D., Van Montagu, M. and Botterman, J. (1990) *Plant Cell*, **2**, 427–436.
29. Buvoli, M., Biamonti, G., Ghetti, A., Riva, S., Bassi, M. T. and Morandi, C. (1988) *Nucleic Acids Res.*, **16**, 3751–3770.
30. Burd, C. G., Swanson, M. S., Gorlach, M. and Dreyfuss, G. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 9788–9792.
31. Bourbon, H., Lapeyre, B. and Amalric, F. (1988) *J. Mol. Biol.*, **200**, 627–638.
32. Cobianchi, F., Karpel, R. L., Williams, K. R., Notario, V. and Wilson, S. H. (1988) *J. Biol. Chem.*, **263**, 1063–1071.
33. Werneke, J. M., Chatfield, J. M. and Ogren, W. L. (1989) *Plant Cell*, **1**, 815–825.
34. Grotewold, E., Athma, P. and Peterson, T. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 4587–4591.
35. Baker, B. S. (1989) *Nature*, **340**, 521–524.
36. Buvoli, M., Cobianchi, F., Bestagno, M. G., Mangiarotti, A., Bassi, M. T., Biamonti, G. and Riva, S. (1990) *EMBO. J.*, **9**, 1229–1235.
37. Ge, H., Zuo, P. and Manley, J. L. (1991) *Cell*, **66**, 373–382.
38. Mayeda, A. and Krainer, A. R. (1992) *Cell*, **68**, 365–375.